# PROBABILITY THEORY FOR DATA SCIENCE

**Prof. Ishapathik Das**

**Department of Mathematics and Statistics**

**Indian Institute of Technology Tirupati**

**Week - 08**

**Lecture - 41**

## Moments for Bivariate Random Variables

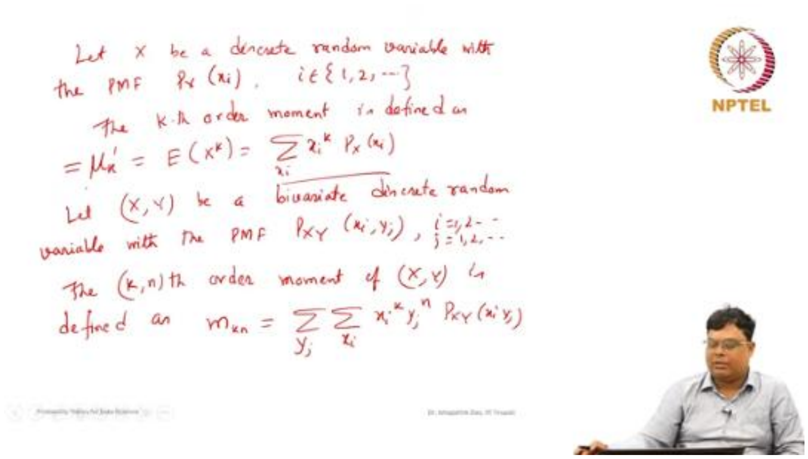
You may recall that when we define the moments for univariate random variables. So, let us discuss. Let X be a discrete random variable with the probability mass function P_X(x_i), where i ∈ {1, 2, ...}. For some range of x, such as $x_1$, $x_2$, $x_3$, and so on, the k-th order moment is defined. As you may recall from our discussion, it is referred to as $\mu^k$, which represents the expected value of x raised to the power of k. Since this is a discrete random variable, it is expressed as the summation of $x^k$.

As you may recall from our discussion, it is referred to as $\mu^k$, which represents the expected value of x raised to the power of k. Since this is a discrete random variable, it is expressed as the summation of $x^k$. The k-th order moment is defined by taking each value $x_i$, raising it to the power of k, multiplying it by the probability of $x_i$, and summing over all $x_i$. If this summation exists, it is referred to as $\mu^k$.

Similarly, for bivariate random variables, the concept is defined as well. Let (X, Y) represent a bivariate discrete random variable with a probability mass function. The probability mass function describes the probabilities for pairs of values, such as $x_i$ and $y_j$.

Here, i takes values like 1, 2, and so on, and j also takes values like 1, 2, and so on. These values can either be finite or countably infinite. In the bivariate case, the (k, n)-th order moment is defined by considering two integers, k and n, which represent the orders of the moments for the two variables. So, the notation for the (k, n)-th order moment is $m_{kn}$, where k and n are positive integers. This is defined as the summation over all possible values of $x_i$ and $y_j$, since it is a discrete random variable. It involves $x_i$ raised to the power of k, $y_j$ raised to the power of n, and the joint probability of $x_i$ and $y_j$.

This is simply an extension to the bivariate random variable cases. For this case, there are some special notations we have already discussed, such as $\mu_x$ for the X random variable and $\mu_Y$ for the Y random variable. Now, let's consider some values for k, such as 1 or 2, and for n, such as 1, 2, and so on. You can also consider 0. So let us consider some of the values. Suppose k is 1, 2, and so on, and n is also 1, 2, and so on.




You can also consider 0. If both k and n are 0, then the result becomes 1. Any value of $x_i$ raised to the power of 0 is 1, and any value of $y_j$ raised to the power of 0 is also 1. This is simply the summation of the probability mass function, which, by the properties of the joint probability mass function, equals 1. Therefore, when both k and n are 0, the result is 1. This is essentially the summation of $x_i$ and $y_j$, multiplied by the joint probability mass function of $x_i$ and $y_j$, which also equals 1 due to the properties of the joint probability mass function.

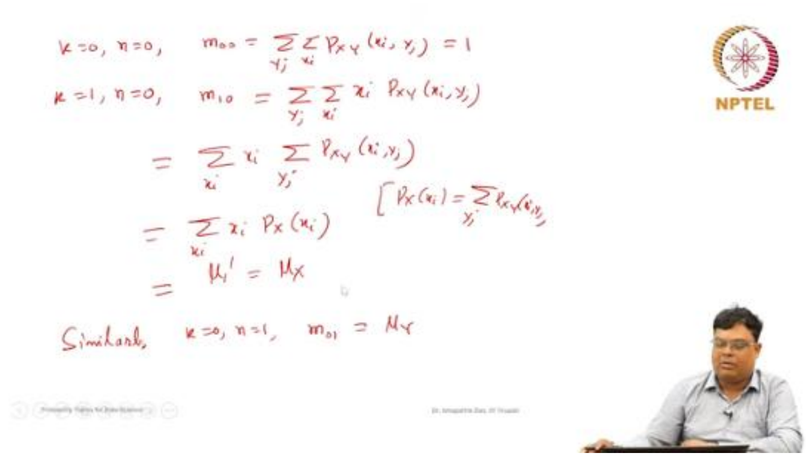
Now, if you consider k as 1 and n as 0, this represents $m_{10}$. What does that mean? It is simply the summation of $y_j$ and the summation of $x_i$. If we apply this with n equal to 0, $y_j$ raised to the power of 0 becomes 1, and $x_i$ is raised to the power of 1. Therefore, this becomes $x_i$ multiplied by the joint probability of $x_i$ and $y_j$.

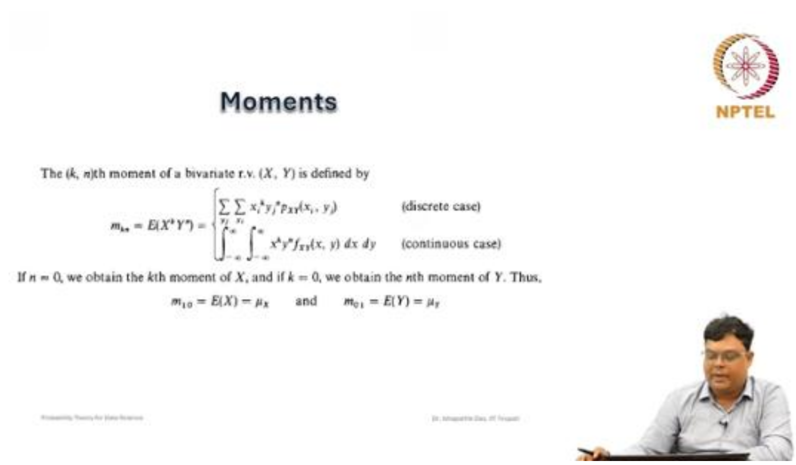
Now, if we assume that the summation exists and we can interchange the summations, let's first consider the sum of $y_j$ and the joint probability of $x_i$ and $y_j$. By definition, this becomes the marginal probability mass function, which is simply the probability mass function of

$x_i$. By definition, you know that to find the marginal probability mass function, you sum over the other variable while fixing $x_i$ for any given value of $x_i$. If you take the sum over the other variable, $y_j$, you get the probability mass function of $x_i$. Then, if you take the sum over $x_i$, you get the probability mass function of $x_i$. This is what you can recall from the definition you just wrote.

So, if you take k as 1, it represents the expected value of X, where X is simply $x_i$. The summation of $x_i$ multiplied by the probability mass function of $x_i$ is nothing but $\mu_1$. Therefore, this is equal to $\mu_x$. That is why, when k is 1, $m_{10}$ is equal to $\mu_x$. Similarly, you can show that $m_{10}$ is also equal to $\mu_x$.

So, whenever you take k as 0 and n as 1, $m_{01}$ represents $\mu_\gamma$. Similarly, you can find this for other values as well. So, we can do it also. What we have discussed so far defines the (k, n)-th moment of a bivariate random variable (X, Y). The (k, n)-th moment is the expected value of X raised to the power of k and Y raised to the power of n.



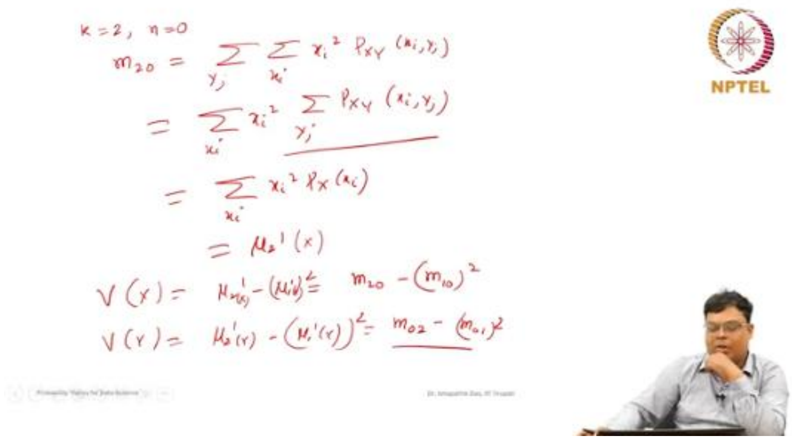
This is the summation of $x_i$ and $y_j$, where $x_i$ is raised to the power of k, $y_j$ is raised to the power of n, and the joint probability mass function of $x_i$ and $y_j$ is considered. So whenever it is a discrete random variable, $m_{10}$ is nothing but $\mu_x$ and $m_{01}$ is nothing but $\mu_y$, so that is already we have discussed. Now, consider the case when k is 2 and n is 0. In this case, it represents $m_{20}$. By definition, $m_{20}$ involves $x_i$ and $y_j$. The (k, n)-th moment is simply the summation over $x_i$ and $y_j$. When n is 0 and k is 2, $y_j$ raised to the power of 0 becomes 1.

This is simply $x_i^2$ multiplied by the joint probability of $x_i$ and $y_j$. Here, we assume that the summation exists and we can interchange the summations. So, this becomes the summation of $x_i^2$ and $y_j$, multiplied by the joint probability mass function. This sum represents the marginal probability mass function of X, which is $x_i^2$ multiplied by the probability mass function of $x_i$. By definition, this is equal to $\mu_2$.

So, $m_{20}$ is equal to $\mu_2$. Therefore, we can find the variance of X, which is $\mu_2 - \mu_1^2$. In this notation, it becomes $m_{20} - m_{10}^2$. This represents the variance of X. Similarly, you can find the same for other variables. In fact, $\mu_2$ refers to the variable X only, and we use the same notation for $\mu_1$.

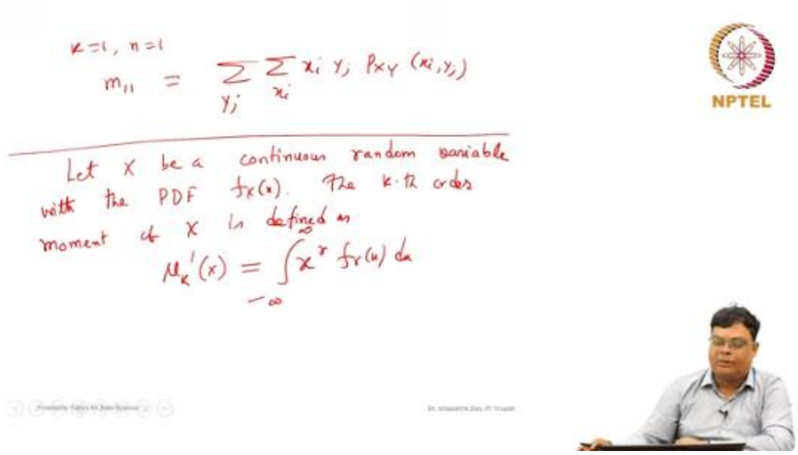
So, $\mu_2$ we can write as the expected value of $X^2$ if you write some notation like for X. Suppose this is X, with $\mu_1$ for X. Similarly, the variance of Y will be $\mu_2$ for Y - $\mu_1$ for Y, all squared. This is equivalent to $m_{02} - m_{01}^2$. Using this method, we can find the means and variances.

So, now what about $m_{11}$? Whenever k is equal to 1 and n is equal to 1, we get $m_{11}$. By definition, this is the summation over $x_i$ and $y_j$, where $x_i$ is raised to the power of 1 and $y_j$ is raised to the power of 1, multiplied by the joint probability of $x_i$ and $y_j$. We may not have a simplified form here because this is specifically for bivariate cases. For univariate cases, this measure does not exist.

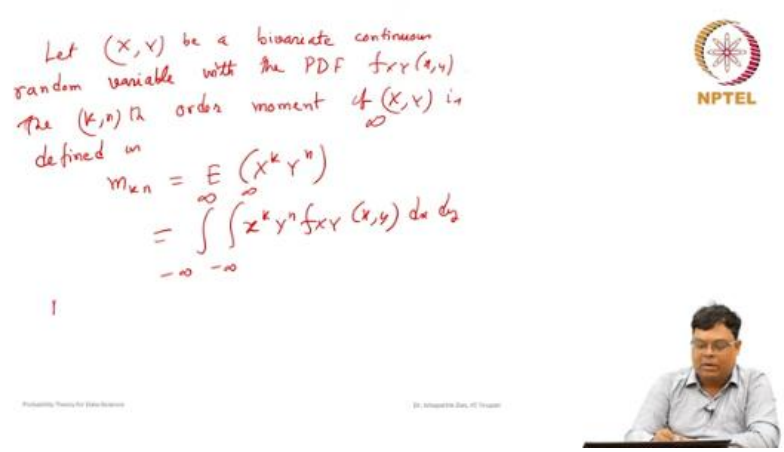



However, this measure provides an indication of the association between $x_i$ and $y_j$, showing how x and y change in relation to their values. We will discuss this later. For now, we will focus on the similar concept for continuous random variables. Now, you can recall that X is a continuous random variable with its probability density function, $f_x(x)$. The k-th order moment has already been discussed in detail. It is referred to as the raw moment. We have also covered central moments, which are moments around a point, but here we are only focusing on the raw moment. For simplicity, the k-th order moment for X is defined and denoted as $\mu_x$ or $\mu_1$. Since there are other variables like X, we write it as (X, $\mu_1$). This is simply equal to... Apologies for the confusion. This is $\mu^r$, not μ. $\mu_x$ refers to the first-order moment, but here we are discussing the k-th order moment.

So, $\mu^r$ is equal to the summation for both discrete and continuous random variables. For a continuous random variable, it is calculated over the range from $-\infty$ to $+\infty$, with X raised to the power r, multiplied by the probability density function $f_x(x)$. Similarly, for a bivariate continuous random variable, the concept will be extended. Let (X, Y) be a bivariate continuous random variable with the probability density function $f_{x\gamma}(x, y)$. The (k, n)-th order moment of (X, Y) is defined as follows.



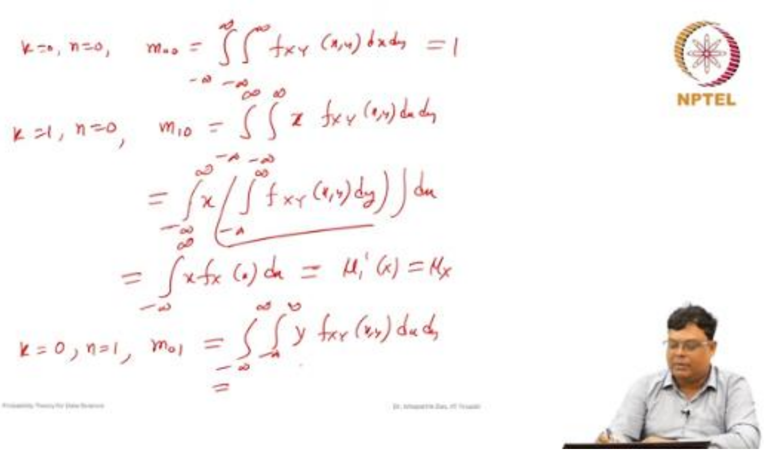

The expected values are denoted as $m_{kn}$, which represent the expected values of X raised to the power k and Y raised to the power n. This is defined as the integral from $-\infty$ to $+\infty$, and from $-\infty$ to $+\infty$, of X raised to the power k, Y raised to the power n, multiplied by the joint probability density function $f_{XY}(x, y)$. This represents the (k, n)-th order moment. Similarly, suppose k is equal to 0, 1, 2, and n is equal to 0, 1, 2, and so on. For $k = 0$ and $n = 0$, we get $m_{00}$. In this case, the integral goes from $-\infty$ to $+\infty$, and since X raised to the power 0 and Y raised to the power 0 equals 1, this becomes the joint probability density function $f_{XY}(x, y)$ integrated over dx and dy.




By the properties of the joint density function, this should be equal to 1. Now, for $k = 1$ and $n = 0$, we get $m_{10}$. In this case, the integral again goes from $-\infty$ to $+\infty$. Since $k = 0$, X raised to the power $k = 1$, and Y raised to the power $0 = 1$, the expression simplifies accordingly. This is X and Y raised to the power 0, multiplied by the joint probability density function $f_{XY}(x, y)$, integrated over dx and dy. Now, assuming that this integration is absolutely summable and exists, we can change the order of integration.

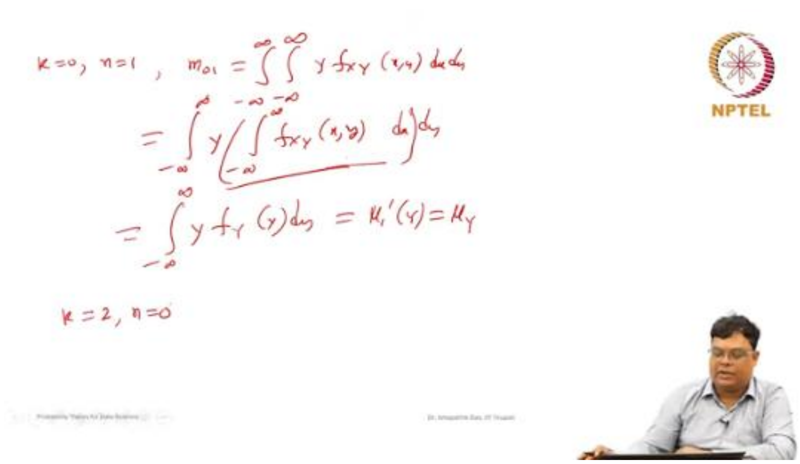
So, the expression becomes the integral from $-\infty$ to $+\infty$ of X, followed by the integral from $-\infty$ to $+\infty$ of $f_{XY}(x, y)$ with respect to y and then with respect to x. Now, what does this represent? You can recall that when we discussed how to find the marginal probability density function from the joint probability density function, this is exactly the joint probability density function. If you integrate with respect to the variable Y, you will get the marginal probability density function, which is $f_X(x)$ dx. Now, if you compare this with the univariate case, suppose $k = 1$. In that case, you get $\mu_1$ for the variable X, which is the integral from $-\infty$ to $+\infty$ of X multiplied by $f_X(x)$ dx. This is the same as the previous value.

This is simply $\mu_1$ for X, or we denote it as $\mu_X$. Similarly, $m_{10}$ is $\mu_1$. Now, when $k = 0$ and $n = 1$, we get $m_{01}$. Similarly, you can find that this is the integral from $-\infty$ to $+\infty$, and from $-\infty$ to $+\infty$, with X raised to the power 0, Y raised to the power 1, and then the joint probability density function $f_{XY}(x, y)$ integrated over dx and dy. If you integrate first with respect to X, you will get the result. For the case where $k = 0$ and $n = 1$, we get $m_{01}$.

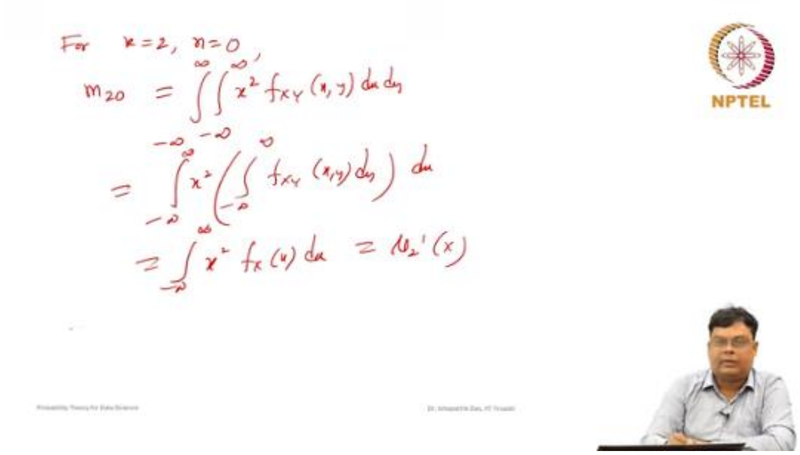



This involves integrating from $-\infty$ to $+\infty$, with X raised to the power 0 and Y raised to the power 1, multiplied by the joint probability density function $f_{XY}(x, y)$, over dx and dy. First, integrating with respect to X gives the marginal probability density function $f_Y(y)$. This simplifies to the integral from $-\infty$ to $+\infty$ of Y multiplied by $f_Y(y)$, over dy. So this is nothing but ∫ from $-\infty$ to $+\infty$ y $f_Y(y)$ dy. This formula is the same and is equivalent to $\mu_1$ for the random variable Y, or simply denoted as $\mu_Y$.

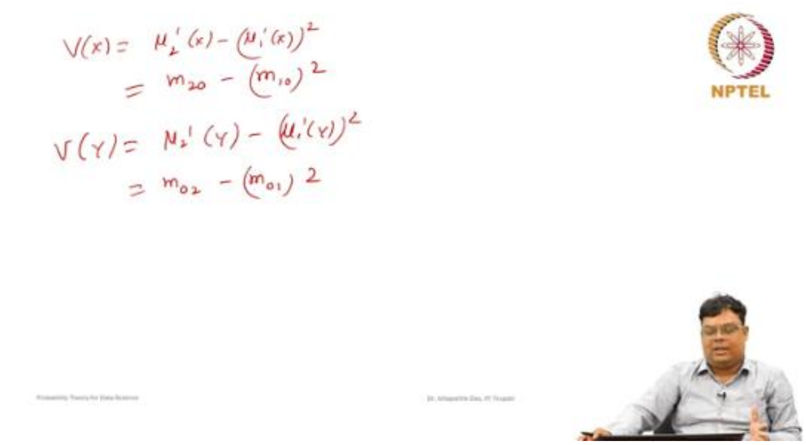
Now, if we take k = 2 and n = 0, let us move to the next case where k = 2 and n = 1. What we get when n = 0 is $m_{20}$. This is represented by X raised to the power 2, which is $X^2$, and Y raised to the power 0, which equals 1, multiplied by the joint probability density function $f_{XY}(x, y)$, integrated over dx and dy from $-\infty$ to $+\infty$. Assuming that the conditions for changing the order of integration are satisfied, we can rewrite it as $X^2$ multiplied by the integral of $f_{XY}(x, y)$ with respect to Y first, followed by integration with respect to X. This is essentially the marginal probability density function of $X^2$.



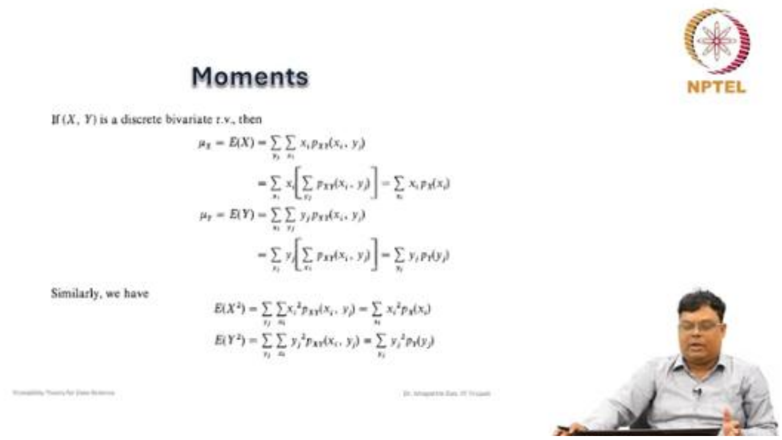

Following that, it represents the probability density function of X multiplied by dx. Ultimately, this corresponds to the second moment about the origin for the random variable X. Similarly, you can find that if you take k = 0 and n = 2, you get $m_{02}$. In the same way, this corresponds to the second moment about the origin for the random variable Y. Hence, the variance of X can be found as $m_{20} - (m_{10})^2$. Similarly, the variance of Y is $m_{02} - (m_{01})^2$.




So, these are some of the values. It is very similar to the discrete cases. These values are derived in the same way. These values are derived in the same way. This is represented in this expression as well. This is the definition, and all the computations are shown here. For the discrete random variables X and Y, we obtain $\mu_x$ and $\mu_\gamma$.

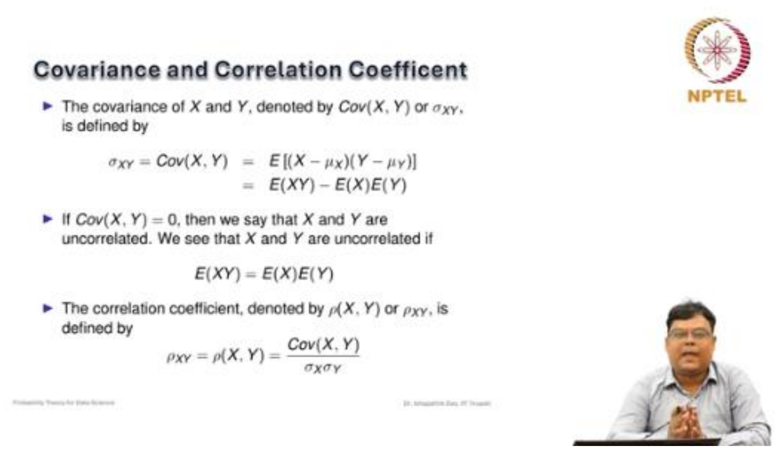



For the discrete random variables X and Y, we obtain $\mu_x$ and $\mu_\gamma$. Similarly, we obtain the expected values of $X^2$ and $Y^2$. For the continuous cases, the values are also provided, and $\mu_x$ and $\mu_\gamma$ are represented in this manner. Likewise, the expected values of $X^2$ and $Y^2$ are given. Now, we will discuss what happens when $k = 1$ and $n = 1$.




Suppose k = 1 and n = 1. For the discrete case, we have already written the expression. For continuous cases, we integrate from -∞ to +∞. This is essentially the product of x and y, multiplied by the joint probability density function f_XY(x, y), integrated over dx and dy. There is no other way to find this value for a particular example, but if it exists, we can calculate it.

Now, we cannot relate this to univariate cases because this measure does not exist for univariate distributions. It is specific to bivariate cases, and in this context, the measure is different. However, this measure is actually significant. We will discuss how we can relate two random variables, suppose X and Y. We have a measure of association that shows how two variables are related to each other.

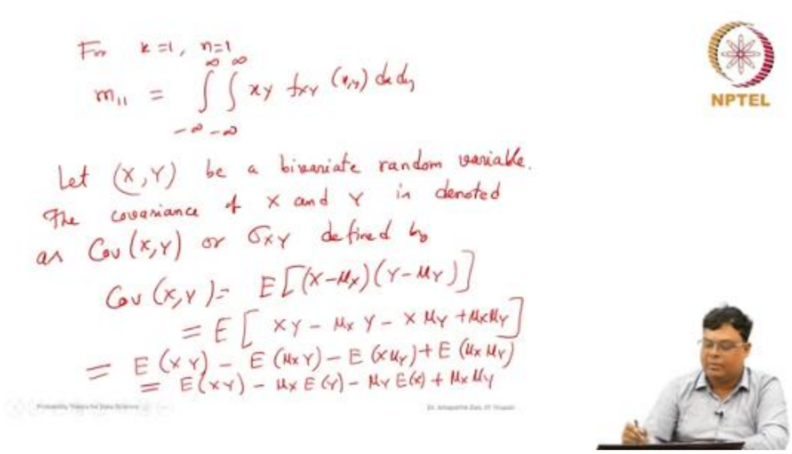



One such measure is called covariance. Let (X, Y) be a bivariate random variable, which could be discrete or continuous. This is the definition of covariance. The covariance of X and Y, denoted as Cov(X, Y) or $\sigma(X, Y)$, is defined by the relation between X and Y. The covariance of X and Y is the expected value of the product of $(X - \mu_X)$ and $(Y - \mu_Y)$.

By subtracting their means and calculating the first-order moments, we get the covariance between X and Y. This can be simplified in a manner similar to variance. So, the expected value of $(X - \mu_X) * (Y - \mu_Y)$ can be written as the product of X and Y, minus $\mu_X * \mu_Y$. Since the expected value is a linear transformation, we know the properties of the expected value. This can be expressed as the expected value of the product of X and Y, minus the expected value of $X * \mu_Y$, minus the expected value of $Y * \mu_X$, plus $\mu_X * \mu_Y$.

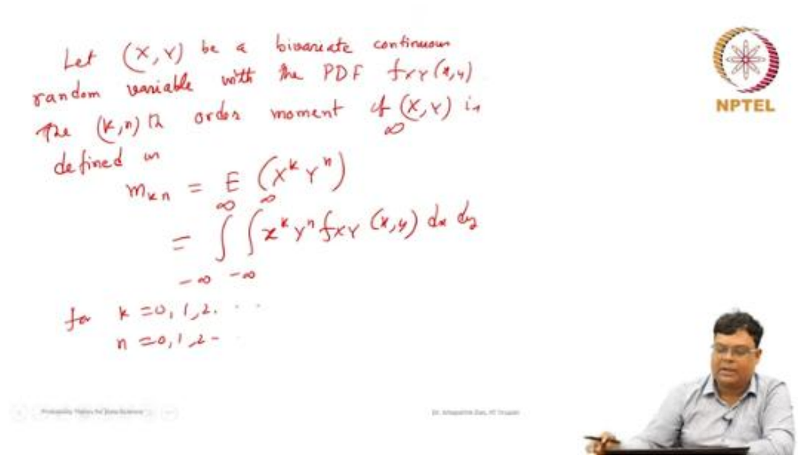
Note that the means of X and Y are constants. So, these are some real numbers after taking the expected value; these are some constants. Then, we will apply the properties of the expected value again. This is the expected value of the product of X and Y, minus the product of $\mu_X$ and the expected value of Y. Since $\mu_X$ is a constant, multiplying a constant by a random variable is equivalent to multiplying the constant by the expected value of that random variable. Similarly, $\mu_Y$ is a constant, so we subtract the product of $\mu_Y$ and the expected value of X.

Finally, we add the product of the means of X and Y, since both are constants. The expected value of a constant is constant, so the product of $\mu_X$ and $\mu_Y$ is constant. Now, if we simplify these terms, we find that this is equal to the expected value of the product of X and Y, minus $\mu_X * \mu_Y$. So, this is the mean of X multiplied by the expected value of Y, minus the mean of Y multiplied by the expected value of X, which is just the notation for the means of X and Y. So, we find that one of the terms involving the means of X and Y cancels out.



Finally, we get the expected value of the product of X and Y, minus the product of the means of X and Y. In our notation, this is represented as the expected value of the product of X and Y, which is denoted as $m_{11}$. This is how we compute the moments. You can see that here. So, when k = 1 and n = 1 in both cases, we compute it from the definition. In this definition, you can see that this is the case.

So, if you set k = 1 and n = 1, this is simply the expected value of the product of X and Y, since m(k, n) is defined as the expected value of X raised to the power of k multiplied by Y raised to the power of n. Finally, $m_{11}$ - $(\mu_x * \mu_y)$ gives us the covariance between X and Y. The covariance provides an association between the two random variables, X and Y, but it is not unitless. Its value depends on the units we are considering. So, if you change the units, the covariance is not bounded, meaning it can take different values.




It can take any value. So, if you want to compare different random variables, suppose X and Y, and another pair, Z and W, and we know the covariance between X and Y, and the covariance between Z and W, we might want to compare how X and Y are related by covariance and how Z and W are related. However, since they depend on different units of measurement, we cannot make a direct comparison. That is why, to make it unitless, a different measure is introduced, called the correlation coefficient. The correlation coefficient is defined as follows.

The correlation coefficient is denoted by ρ(X, Y) and is defined as ρ(X, Y) or sometimes written as ρ(X, Y) with parentheses. This notation is used, and it is defined by the covariance of X and Y divided by the product of the standard deviations of X and Y. We can express it as $\sigma(X, Y) / \sqrt{\sigma^2(X) * \sigma^2(Y)}$. The covariance of X and Y is represented by $\sigma(X, Y)$, while $\sigma^2(X)$ represents the variance of X, and $\sigma^2(Y)$ represents the variance of Y. Next, we will discuss some properties of covariance, how they are associated, and the properties of the correlation coefficient.

Based on the values of the correlation coefficient, we can determine how X and Y are related. We will also explore the properties of the correlation coefficient in the upcoming discussion.